TRANSLATION

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY
(Chapter II of the Patent Cooperation Treaty)

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference **2003ODT0902P**	**FOR FURTHER ACTION**	See Form PCT/IPEA/416
International application No. **PCT/EP2004/009539**	International filing date *(day/month/year)* **26.08.2004**	Priority date *(day/month/year)* **30.09.2003**

International Patent Classification (IPC) or national classification and IPC

G06K9/20

Applicant

OCE DOCUMENT TECHNOLOGIES GMBH

1. This report is the international preliminary examination report, established by this International Preliminary Examining Authority under Article 35 and transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of **15** sheets, including this cover sheet.

3. This report is also accompanied by ANNEXES, comprising:

   a. ☒ *(sent to the applicant and to the International Bureau)* a total of **6** sheets, as follows:

      ☒ sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications authorized by this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions).

      ☐ sheets which supersede earlier sheets, but which this Authority considers contain an amendment that goes beyond the disclosure in the international application as filed, as indicated in item 4 of Box No. I and the Supplemental Box.

   b. ☐ *(sent to the International Bureau only)* a total of (indicate type and number of electronic carrier(s))

   _____ , containing a sequence listing and/or tables related thereto, in computer readable form only, as indicated in the Supplemental Box Relating to Sequence Listing (see Section 802 of the Administrative Instructions).

4. This report contains indications relating to the following items:

   ☒ Box No. I   Basis of the report

   ☐ Box No. II   Priority

   ☐ Box No. III   Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

   ☐ Box No. IV   Lack of unity of invention

   ☒ Box No. V   Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

   ☐ Box No. VI   Certain documents cited

   ☐ Box No. VII   Certain defects in the international application

   ☐ Box No. VIII   Certain observations on the international application

Date of submission of the demand	Date of completion of this report
Name and mailing address of the IPEA/EP	Authorized officer
Facsimile No.	Telephone No.

Form PCT/IPEA/409 (cover sheet) (January 2004)

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.

PCT/EP2004/009539

Box No. I	Basis of the report

1. With regard to the language, this report is based on the international application in the language in which it was filed, unless otherwise indicated under this item.

   ☐ This report is based on translations from the original language into the following language _____ , which is the language of a translation furnished for the purposes of:

   ☐ international search (Rule 12.3 and 23.1(b))

   ☐ publication of the international application (Rule 12.4)

   ☐ international preliminary examination (Rule 55.2 and/or 55.3)

2. With regard to the elements of the international application, this report is based on *(replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report)*:

   ☐ the international application as originally filed/furnished

   ☒ the description:

       pages   1-18 _____ as originally filed/furnished

       pages*   3a,3b _____ received by this Authority on   11.07.2005 with letter of 11.07.2005

       pages* _____ received by this Authority on _____

   ☒ the claims:

       nos. _____ as originally filed/furnished

       nos.* _____ as amended (together with any statement) under Article 19

       nos.*   1-19 _____ received by this Authority on   11.07.2005 with letter of 11.07.2005

       nos.* _____ received by this Authority on _____

   ☒ the drawings:

       sheets   1/8-8/8 _____ as originally filed/furnished

       sheets* _____ received by this Authority on _____

       sheets* _____ received by this Authority on _____

   ☐ a sequence listing and/or any related table(s) – see Supplemental Box Relating to Sequence Listing.

3. ☐ The amendments have resulted in the cancellation of:

   ☐ the description, pages _____

   ☐ the claims, nos. _____

   ☐ the drawings, sheets/figs _____

   ☐ the sequence listing *(specify)*: _____

   ☐ any table(s) related to sequence listing *(specify)*: _____

4. ☐ This report has been established as if (some of) the amendments annexed to this report and listed below had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).

   ☐ the description, pages _____

   ☐ the claims, nos. _____

   ☐ the drawings, sheets/figs _____

   ☐ the sequence listing *(specify)*: _____

   ☐ any table(s) related to sequence listing *(specify)*: _____

* *If item 4 applies, some or all of those sheets may be marked "superseded."*

Form PCT/IPEA/409 (Box No. I) (January 2004)

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1.   Statement

    Novelty (N)   Claims 1-19 _____ YES

                       Claims _____ NO

    Inventive step (IS)   Claims _____ YES

                       Claims 1-19 _____ NO

    Industrial applicability (IA) Claims 1-19 _____ YES

                       Claims _____ NO

2.   Citations and explanations (Rule 70.7)

1   This report makes reference to the following documents:

      D1:   US 5 317 646 A (SANG JR HENRY W ET AL) 31 May 1994 (1994-05-31)

      D2:   US 2002/141660 A1 (PUCCI JORGE PABLO ET AL) 3 October 2002 (2002-10-03)

      D3:   US 6 028 970 A (DIPIAZZA PHILIP SILVANO ET AL) 22 February 2000 (2000-02-22)

2   The subject matter of claim 1 fails to involve an inventive step (PCT Article 33(3)).

2.1   Document D2 is considered the closest prior art and discloses (the references between parentheses refer to that document; passages that are struck through (~~such as here~~) indicate passages from claim 1 which have no equivalent in D2):

        a method for acquiring data from machine-readable documents, ~~the data being allocated to a database,~~ in which individual data items are extracted from the document and entered into corresponding

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

~~database~~ fields in as fully automated a manner as possible,

(abstract: "The document scanner, system and , method operates in conjunction with a document imprinted with data and a plurality of form documents adapted to have data imprinted thereon. The documents have at least one and typically many data image fields. Ultimately, the document scanner, system and method output a delimited string of decoded characters to another computer system via a common computer communications port. ... The system selects one of the stored forms, extracts the data from each data field, decodes or calculates the data, and validates the data (in the presence of data validation parameters) and stores the decoded/calculated data.", end of paragraph [0052]: "It should be appreciated that the further computer device can easily process this delimited string of decoded characters into a spreadsheet, database or any other type of word processing program.")

and if data for one or more specific ~~database~~ fields cannot be extracted from the document with the necessary level of reliability

(end of abstract: "A data reporting and data correction system, activated in the presence of the data error reporting and correction descriptor, enables correction of errors"),

the following steps are executed:
- displaying the document on screen (implicit),
- displaying on screen the ~~database~~ field for

which the data cannot be extracted with the
necessary level of reliability,

(paragraph [0051]: "Any error reports from field
and rule checker unit 62 are supplied via control
unit 80 to display 84. The operator at keyboard
86 may correct the error if the data correction
field descriptor has been turn[ed] ON. If the
operator is enabled to correct the data and does
correct the error, summation module 64 substitutes
the corrected data for the previously scanned and
decoded incorrect data.")

- ~~execution of a proposal routine, with which
string sections in the vicinity of a cursor
that can be moved on the screen by a user are
selected, marked and proposed for extraction.~~

2.2   The subject matter of claim 1 differs from the
teaching of D2 by virtue of the following points:

i)   the data to be extracted is allocated to a
database;

ii)   the acquired data is entered into the
database fields;

iii) in cases where (database) fields cannot be
extracted with the necessary level of
reliability, a proposal routine is executed
with which string sections in the vicinity of
a cursor that can be moved on the screen by a
user are selected, marked and proposed for
extraction.

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

2.3   The stated differences allow a number of interpretations. The differences were interpreted as follows (numbers i)-iii) correspond to the numbers indicated above):

i)   an option is provided for storing the data in a database, i.e. a database and a list of database fields together with correspondences to fields that are extracted from the document are known from the method;

ii)   the acquired data is stored in components of a data structure which must be used when the data is to be stored in a database, i.e. ii) is implicit from i).

iii)   a routine for controlling a "mouse" cursor which allows string sections to be selected, for example by defining a rectangular section of the screen, is also regarded as a "proposal routine with which string sections in the vicinity of a cursor that can be moved on the screen by a user are selected, marked and proposed for extraction".

It is the examiner's opinion that the chosen wording does not indicate that the proposal routine uses recognised (alphanumerical) data (and uses only the position thereof in the bitmap) and does not, as in D2, define image sections based on a pixel-type procedure or extract string sections using OCR.

This opinion is *in addition* supported by the

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

wording of claim 8, according to which "and the proposal routine presents, *in addition* to the graphic representation of the marked string section, the coded text of that string section".

2.4 The present invention can therefore be considered to address the following problems:

i) and

ii) making it possible to store the extracted data in a database (i.e. of a structured, durable, searchable storage system).

iii) devising a convenient way of inputting corrections in the event of errors or uncertain results during the data extraction.

2.5 The differences or problems specified under i) and ii) on the one hand and iii) on the other hand are completely independent of each other and thus the pertaining features represent a juxtaposition of features.

2.6 Therefore, in assessing the involvement of an inventive step, i) and ii) on the one hand and iii) on the other hand are considered independently of and separately from one another.

2.7 Regarding i) and ii): the possibility of storing the data in a database is already considered in D1 (end of paragraph [0051]: "It should be appreciated that the further computer device can

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

easily process this delimited string of decoded characters into a spreadsheet, database ...").
Sections [0036]-[0041] show that in the method as per D2 the field information (metadata) needed for connecting to the database is provided. Since claim 1 does not contain any further information relating to the connection to the database, the stated passage from D2 is taken as sufficient indication for a person skilled in the art to be able to derive aspects i) and ii) of claim 1 from D2.

Irrespective thereof, documents D1 (fig. 2, 700: "Database Insertion") and D3 (fig. 1B) show that the storage of data extracted from documents in a database is known.

Regarding iii): as is shown in paragraphs [0050] and [0051] of D2, the extracted data is checked. If an error is found, manual correction is possible. D2 does not provide exact details regarding the manual correction.

A person skilled in the art charged with implementing a manual error correction method that is simple and convenient for the user would recognise that the following must be shown:
- for which field invalid data has been extracted;
- from where the data stems (in the bitmap of the scanned-in document).

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

It is recognised that a person skilled in the art could consider showing only the erroneous data and the section of the bitmap from which the erroneous data has been extracted. The user of the method must then clearly manually enter the data via the keyboard. It is, however, considered that a person skilled in the art would most likely consider devising a simple option with which a section or sections of the bitmap of the document is or are selected and the already present OCR function is used to extract the data from the bitmap.

The required functionality thus corresponds to the functionality needed for field definition (and therefore can be at least partly re-used). According to section [0042] of D2, a cursor is used to define the position and size of the fields.

It is therefore considered that the features as per point iii) are obvious to a person skilled in the art from the teaching of D2 alone.

Irrespective thereof, a person skilled in the art is familiar from D1 with a method which describes a particularly simple definition of parts of a document as fields to be extracted. In contrast to the method suggested by D2, it is not necessary in the D1 method to manually define the size of the section to be extracted.

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

3   Independent claim 12[1] does not meet the requirements of PCT Article 6. The subject matter of claim 12 appears not to involve an inventive step (PCT Article 33(3)).

3.1   The passage in lines 20-27 of claim 12 is understood to mean that a comparison of *the content* of a string section which appears below in a table with string sections which appear in the first few lines is used to determine from which field the string section must be extracted.

Such a method, however, delivers the desired result only in exceptional cases. To that end, the columns must contain almost identical entries and the columns must differ significantly from one another. It is therefore not clear to the reader what is the intended scope of protection.

Furthermore, the indicated interpretation is inconsistent with the description. The extraction of data from tables is described on page 12, line 1 - page 15, line 10; the passage on page 12, line 31 - page 13, line 7 describes the comparing of string sections using a cost function.

---

[1] Owing to the use of "in particular" in the clause "in particular according to one of claims 1 - 11", claim 12 cannot be considered to be *necessarily* dependent on one of claims 1 - 11.

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

According to the latter passage, the *horizontal position and the width* of the string sections are compared.

The wording chosen for claim 12 is thus considered to be misleading and to not be supported by the description.

3.2   Document D3 indicates that rules are established which describe the data lists or tables and are used for extraction and error analysis and correction (column 2, lines 38-45: "A third type of rule is a position verifier. This type of rule requires that certain ordering logic inherent in the definition of the data fields be followed …", column 13, lines 36-40, fig. 4, fig. 5). Manual correction is provided for; see figs. 1B, 1C: "Operator review, if required".

The subject matter of claim 12, as it is understood in the light of the description, is therefore considered to be suggested by D3.

3.3   The subject matter of claim 12 cannot be considered inventive when the wording "in particular" is removed from the claim (in which case claim 12 would be dependent on claim 1), since, as outlined in section 2 above, document D2 renders obvious the subject matter of claim 1 and the features of claim 12 that are not known from D2 are independent (in the sense of a juxtaposition) from the features of claim 1 that

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

are known from D2. The features of claim 12 that are not known from D2 are rendered obvious by D3.

4   The subject matter of claim 16 does not involve an inventive step (PCT Article 33(3)) because the subject matter of claims 1 and 12 does not involve an inventive step and the hardware components used as per claim 16 are common.

5   The subject matter of claim 19 does not involve an inventive step (PCT Article 33(3)) because the subject matter of claims 1 and 12 does not involve an inventive step and because it is assumed that the method from document D2 is implemented in the form of a computer programme product.

6   DEPENDENT CLAIMS 2-11, 13-15, 17, 18

Claims 2-11, 13-15, 17 and 18 do not contain any features which, in combination with the features of any claim to which they refer, meet the PCT requirements for novelty and inventive step.

6.1   The subject matter of claims 2, 3, 4 and 5 is suggested by D2 (fig. 1: "Dictionaries (64)", "Logical Check (68)", "Fields & Interfield Rule Checker (62)"). It is also indicated that the term "concept information" is not normally used as a preamble for syntactic and semantic information. Claim 4 is therefore unclear (PCT Article 6).

6.2   The additional features of claims 6, 7 and 9-11

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

are insignificant.

6.3   Claim 8: document D2 states that discovered errors
       can be manually corrected; see paragraph [0051]:
       "Any error reports from field and rule checker
       unit 62 are supplied via control unit 80 to
       display 84. The operator at keyboard 86 may
       correct the error if the data correction field
       descriptor has been turned ON.". In order for the
       user to be able to make the correction, the
       recognised text must also be displayed.

       The additional features in claim 8 are therefore
       considered to be suggested by D2.

6.4   Claim 13
i.   The phrase "string matching method" describes a
       large class of methods. The phrase is frequently
       used to describe methods for adapting
       (alphanumerical) character strings. Although in
       the method described in claim 13 alphanumerical
       character strings are processed, the finding of
       (partial) matches is based not on alphanumerical
       characters, but on rectangular screen sections,
       the similarities of which are determined using the
       extent of the matching of the positions and sizes
       (and not using the character sequences previously
       found in those sections). The phrase "string
       matching method" is therefore misleading and
       claim 13 is unclear (PCT Article 6).

       The passage in lines 4-9 on page 14 of the

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

description of the current application is very much consistent with the usual meaning of "string matching method". It relates, however, to the testing of the consistency of a string section with concept information and not to a comparison of two string sections that have been extracted from the document. If it were intended that the "string matching method" refer to that passage of the description, then much clearer wording should have been chosen.

ii. Claim 13 does not specify in what way the string matching method is used in the defined method.

iii. Document D3 indicates that rules are used to position fields and that the rules can be combined. This is regarded as a type of string matching method (in the sense that it is used in the description: comparison of the position and/or width of image sections representing the string sections). The additional features in claim 13 are thus known from D3.

6.5 The additional feature in claim 14 is insignificant.

6.6 As is indicated under point 3.2 above, D3 discloses a possibility of manual correction. This is equivalent to the "editing functions" specified in claim 15 and therefore the subject matter of claim 15 is suggested by D3.

Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

6.7   The additional features in claims 17 and 18 are

insignificant and are also known from D2.